Mapping Relational Efficiency in Neuro-Fuzzy Hybrid Cost Models

Citation for published version:
Tokede, O, Ahiaga-Dagbui, D, Smith, S & Wamuziri, S 2014, Mapping Relational Efficiency in Neuro-Fuzzy Hybrid Cost Models. in 2014 Construction Research Congress. 2014 Construction Research Congress, Georgia, United States, 19/05/14.

Published In:
2014 Construction Research Congress

Mapping Relational Efficiency in Neuro-Fuzzy Hybrid Cost Models

Olubukola TOKEDE, S.M.ASCE[1]; Dominic AHIAGA-DAGBUI[2]; Simon SMITH[2] and Sam WAMUZIRI, M.ASCE[3]

[1] School of Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, EH10 5DT, UKPH (44) 131-455-2837; email: O.Tokede@napier.ac.uk

[2] School of Engineering, University of Edinburgh, EH9 3JL, UK; email: D.Ahiaga-Dabui@ed.ac.uk, S.Smith@ed.ac.uk

[3] Centre for Management Research, Glyndŵr University, Wrexham, LL11 2AW, UK, S.Wamuziri@glyndwr.ac.uk

ABSTRACT

Significant improvements are achievable in the accuracy of cost estimates if cost models adequately incorporate issues of flexibility and uncertainty. This study evaluates the relational efficiencies of the fuzzy composition operators – the max-min and max-product, in establishing the final cost of water infrastructure projects. Cost and project data was collected on 1600 water infrastructure projects completed in the UK between 2000 and 2011. Neural network is first used to develop relative weightings of relevant cost predictors. These were then standardized into fuzzy sets to establish a consistent effect of each variable on the overall target cost. The strength and degree of relationship of the normalized cost predictor weightings and the fuzzified project attributes were combined using the max-min and max-product composition operators to obtain project cost predictions. The predictions from the two composition operators are compared with the actual cost figures. Results show comparable performance in the efficiency of the composition operators. Based on statistical correlations, the max-product composition operator achieved on average a deviation of 1.71% while the max-min composition had an average deviation of 1.86%. Improvements in the relational efficiency of neuro-fuzzy hybrid cost models could assist in developing a robust framework for realistic cost targets on construction projects.

INTRODUCTION

One of the major challenges of forecasting is dealing with uncertainty (Hüllermeier 1997) - the broad range of variability of likely outcomes of any event. One approach to uncertainty analysis that allows for some degree of flexibility is the fuzzy sets framework. To a reasonable extent, fuzzy sets basically imply the inclusion of degree of belonging in evaluating variables (Zadeh, 2008). They help to capture irreducible uncertainty as well as model vagueness in human reasoning abilities. Fuzzy relations are special cases of fuzzy sets. Fuzzy relations can be defined as a vague relationship between some fixed numbers of variables (Chan *et al.,* 2009; Zimmerman, 2001). Relations in this case are normative structures that help to interpret the attributes of fuzzy systems. The composition operation is however one class of similarity relation that seeks to establish the relationship between similar elements in different universe of discourse (Zimmermann, 2001). Two common forms of composition operations are the max–product and max–min compositions. Zimmerman (2001) opines that the max – min composition is the most frequently used and that the operations of fuzzy relations can be well defined using the Extension principle. This paper provides an evaluation of the max-min and max-product composition operator in neuro-fuzzy hybrid cost models. The paper briefly discusses construction cost estimation and neuro-fuzzy modelling before detailing the mapping strategies in neuro-fuzzy hybrid cost models. The paper then proceeds to evaluate the relational efficiencies of two composition operators in a neuro-fuzzy hybrid cost estimation model and concluding with results achieved and their implications for research using the two mapping strategies.

COST ESTIMATION

Effective cost estimation relates the design of constructed facilities to their cost, so that while taking full account of quality, risks, likely scope changes, utility and appearance, the cost of a project is planned to be within the economic limit of expenditure (Kirkham and Brandon 2007). This stage in a project life-cycle is particularly crucial as decisions made during the early stages of the development process carry more far-reaching economic consequences than the relatively limited decisions which can be made later in the process. As noted by Hegazy (2002), in spite of the importance of cost estimation, it is undeniably neither simple nor straightforward because of the lack of information in the early stages of the project. Cost estimation is so vital; it can seal a project's financial fate (Nicholas 2004). Rightly, or wrongly, cost estimates produced at the beginning of a project are used by the client to build their budget which often becomes 'the baseline' on which actual project performance may be measured and compared.

Cost estimation techniques range from model-based methods to model-free methods. In between these spectra lie a variety of techniques available to estimate the cost of a project including traditional bills of quantity, activity schedule and detailed estimation. Model-based techniques consist of static sets of relationships, which systematically handle inputs and methodologically translate them into output (Smit 2012). In situations where such relationships are analytical, they mimic some form of mathematical function (Ross 2009). Model-free techniques are more dynamic and adaptive and include fuzzy systems and neural networks (Lee & Lin, 1992).

NEURO-FUZZY COST MODELS

Artificial Neural Networks (ANN), henceforth referred to as neural networks (NN) with artificial implied, is an analogy-based, non-parametric information-processing system that has performance characteristics similar to a biological neural network of the brain (Anderson and McNeill 1992). They retain two features of the biological neural network: the ability to learn from experience and make generalisations based on this acquired knowledge (Haykin 1994). Neural networks are structured to provide the capability to solve problems without the benefits of an expert and without recourse to programming (Boussabaine and Elhag 1999)

Neural networks are promising tools when used in conjunction with fuzzy sets for developing adaptive systems (Kosko and Isaka 1993). Adaptive systems can generally identify rule patterns in incoming data. Neural network and fuzzy logic systems are both numeric model-free estimators and dynamic systems (Lee and Lin 1992). Neural networks provide a platform for classifying patterns without having to provide explanations on the possible sophistications employed by the classification machinery (Eklund 1994). The disadvantage in the neural network technique is that they often increase nodes sporadically or swap network structure arbitrarily (Lee and Lin 1992); a variability that puts to question its reliability. Besides, the blackbox-ness of neural networks, more or less consigns it to the realm of magical arts. Fuzzy models, on the other hand deteriorate significantly where data sets used for identification are highly heterogeneous (Pedrycz 1996). Moreso, its procedures do not seem easily understandable to many cost and construction professionals (Tokede and Wamuziri 2012). Synergizing neural network and fuzzy systems therefore provides promising potentials for intelligent hybrid systems (Lee and Lin 1992). Lin and Lee (1992) pointed out that hybrid learning algorithms perform better than supervised learning algorithm alone. In a more recent study by Ahiaga-Dagbui and Smith (2012), it was discovered that the best neural network models for 98 water infrastructure projects had an average underestimation and overestimation of 1.2% and 4.6% respectively. In comparison, the neuro-fuzzy hybrid cost model using the same dataset achieved an average performance of 0.6% and 0.8% (Ahiaga-Dagbui et al. 2013). Neuro-fuzzy techniques are one of the most common hybrid techniques employed in cost estimation problems. According to Chan et al (2009), such techniques are highly competent in handling pattern recognition and automatic learning. Ahiaga-Dagbui *et al.,* (2013) also suggest that fuzzy sets and neural networks both provide excellent mapping interphases which when combined could be invaluable in pattern recognition.

Mapping Strategies in Neuro-Fuzzy Cost Models

Fuzzy sets are useful in mapping non-empty sets to partially ordered sets (Sanchez 1976). They can be used to bridge the gap between mathematical models and their associated physical reality (Demicco and Klir 2003). This is mainly achieved by representing the vagueness associated with the linguistic description. Fuzzy relations are essentially the means of modelling the intensity between elements of a fuzzy set. Fuzzy relations emerge from Cartesian representation of two or more sets on a universal scale (Belohlavek and Klir 2011).

A composition is a common mathematical operation that seeks to establish the relationships between similar elements in different universe of discourse (Zimmermann 2001). The compositionality assumption is a sort of logical generalization presupposing that the degree of membership of a compound fuzzy set is a function of the membership degrees of each component. Effectively, this implies the whole is summarily a sum and/or product of its parts (Belohlavek and Klir 2011). There have been contention on the possibility of a single non-parametric operator to appropriately model the meaning of ‘AND’ or ‘OR’ context independently. The composition method is commonly used in applications of artificial neural network for mapping between parallel layers in a multi-layer network.

According to Ross (2009), the fuzzy relation, $\tilde{T}$ of two sets, $\tilde{R}$and $\hat{S}$ can be defined by the set-theoretic and membership function-theoretic, mathematically expressed as:

$$\tilde{T}=\tilde{R} \circ \tilde{S} \qquad \text{Eqn. 1}$$

Where R is a fuzzy relation on the Cartesian space X x Y. S is a fuzzy relation on Y x Z, and T is fuzzy relation on X x Z. In this cost estimation problem, R represents the set of cost predictors and S refers to the set of standard values of tolerance for linguistic descriptors of project attribute

Max-min Composition

The max-min composition is commonly used when a system requires a conservative solution. Loetamonphong and Fang (2001, pp6) explains this approach as when the "goodness of one value cannot compensate the badness of another value". Figure 1 shows a graphical illustration of the max-min composition. Ross (2009) pointed out the max-min composition is analogous to approximate reasoning using the IF-THEN rules.

Mathematically, the max-min composition can be represented as:

$$\tilde{\mu}_{T(x,z)} = \vee_{y \in Y}[\tilde{\mu}_{R(x,y)} \wedge \tilde{\mu}_{S(y,z)}] \qquad \text{Eqn. 2}$$

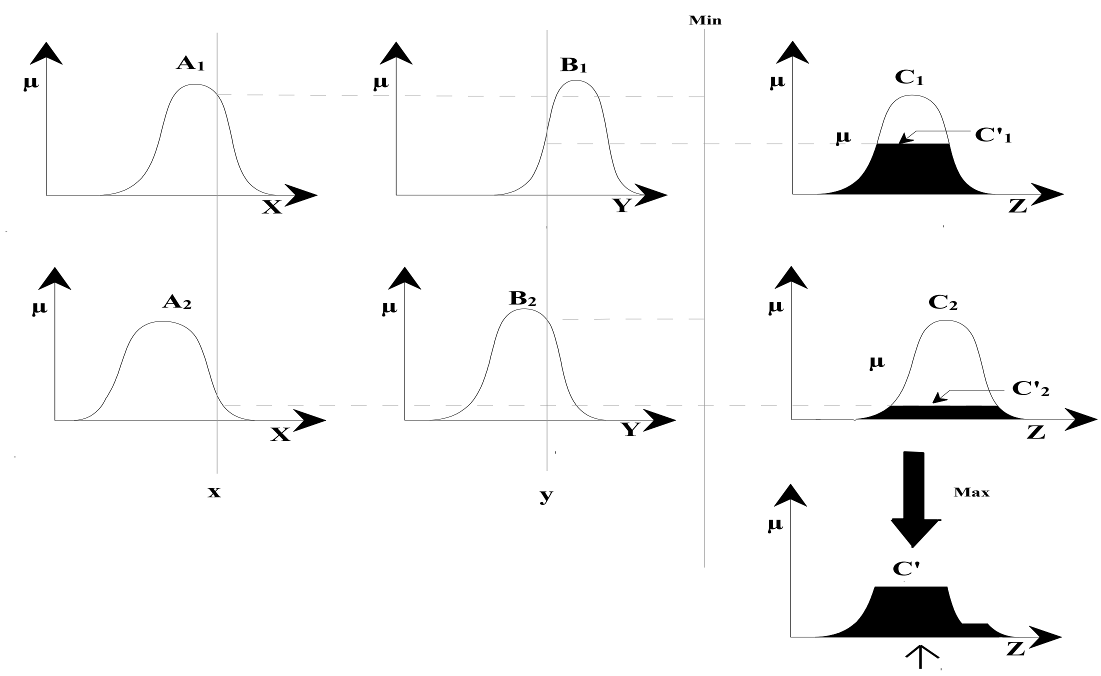

Figure 1 – Graphical illustration of the max-min composition (Dubois & Prade, 2000)

Max-Product Composition

The max-product composition is touted by some researchers as yielding better equivalent results (Loetamonphong and Fang 2001; Ross 2009). One possible explanation is that conventional risk calculus is presumed to have a combinatorial character.

Mathematically, the max-product composition can be represented as:

$$\tilde{\mu}_{T(x,z)} = \vee_{y \in Y}[\tilde{\mu}_{R(x,y)} \bullet \tilde{\mu}_{S(y,z)}] \qquad \text{Eqn. 3}$$

The max-product composition is a fuzzy calculus that expresses the relationship between similar elements. Figure 2 shows a graphical illustration of the max-product composition. Ross (2009) illustrated the max-product composition to relate the rain gauge prediction of large storms to the actual pond performance during rain events.

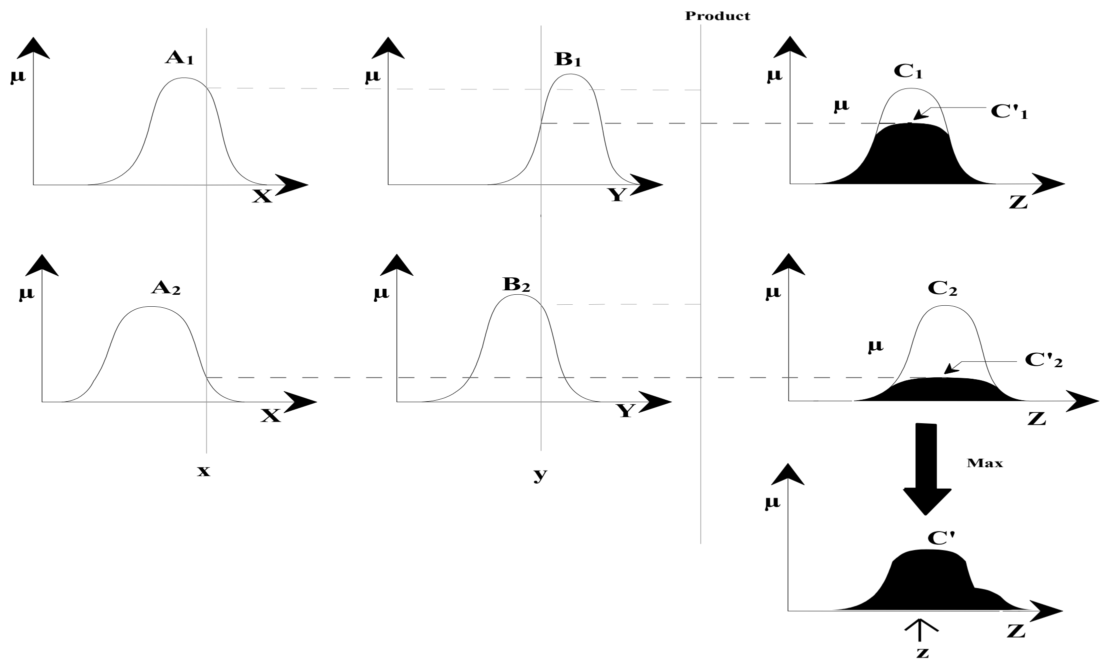

Figure 2 – Graphical illustration of the max-product composition (Dubois & Prade, 2000)

Other possible variants of composition include the max-max, min-min, max-average and sum-product (Ross 2009). Essentially, the composition involves employing hybrid formulations of min, max, average and product to arrive at some relationship formation; thereby specifying a range of mathematical values that could be tolerated by a category (Carpenter *et al.* 1992). Yager and Filev (1994) mentioned that the MAX operator ignores reinforcement inherent in the overlapping in the output fuzzy sets. Carpenter *et al.,* (1992) also stated that the MIN operator helps highlight features that are critically present, whilst the MAX operator flags-off features that are critically absent.

RESEARCH METHOD

The findings reported in this experimental paper were achieved using the following steps. Approximately 1600 projects completed between 2004 and 2012, with cost range of between £4000 to £15 million, comprising newly built, upgrade, repair or refurbishment projects were used for the study. One hundred cases were selected using stratified random sampling to be used for independent testing of the final models. The remaining data were then split in an 80:20% ratio for training and testing of the neural network model. All cost values were normalized to a 2012 baseline with base year 2000 using the infrastructure resources cost indices by the Building Cost Information Services (BCIS 2012). The nature of the projects ranged from construction of water mains, water treatment plants, Combined Sewer Overflows (CSOs), installation of manholes or water pumps and upgrades and repairs to sewers.

The data was then pre-processed to structure and present the data to the model in the most suitable way. For this research, extreme values and outliers were either re-coded or deleted from the sample set and missing values replaced with the mean or mode. Input errors were corrected and all cost values were normalized to 2011 with the base year 1995 using the infrastructure resources cost indices by the Building Cost Information Services (BCIS 2012). Invariant variables, such as procurement option, payment method, fluctuation measure and type of client, were removed from the variable set as they would only increase the model complexity whiles offering little to no useful information for model's performance. Categorical variables such as type of project, need for project, etc. were coded using a binary coding (0, 1) format. Data screening using scree test and optimal binning allowed for the selection of five initial predictors (primary purpose of project, project scope, project delivery partners, estimated target cost and project duration) to be used for the actual ANN modelling. Several neural network models were then developed with the 20 best models used to estimate the relative contribution to model performance of each factor used. These values, as shown in Table 1 were then standardized into fuzzy sets in the next phase of the study to establish a consistent effect of each variable on the overall target cost.

Fuzzy Sets Modelling

Fuzzy set theory is applied at this stage of the modelling exercise to evaluate the subjective measures for each of the cost predictors in order to predict final cost. Using Eqn.4 the average weighted ranking for each of the variables from Table 1 was normalized to unity in order to generate a standardised index for the subsequent fuzzy set computations (see Table 2)

$$\sum Normalized\ ranking = \frac{w_i}{\sum w} = 1 \qquad \text{Eqn. 4}$$

Where w_i is the average relative weighting of the ith predictor

$\sum W$ is the sum of relative weighting of all predictors

Table 1 - Normalized weighted values of the cost predictors from the neural network analysis

Factors	Project Scope	Primary Purpose	Delivery Partner	Duration	Target Cost
Normalized ranking	0.22	0.11	0.02	0.02	0.63

With mean target cost to predictor plots, all predictors were fuzzified using the range set below:

$outturn\ cost \geq £600{,}000$, Influence is Rather High

$£400{,}000 \geq outturn\ cost \geq £600{,}000$ Influence is High

$£100{,}000 \geq outturn\ cost \geq £400{,}000$ Influence is Medium

$outurn\ cost \leq £100{,}000$, Influence is Low

The next stage of the fuzzy modelling involved developing membership functions. In developing these, the tolerance index is particularly relevant in evaluating and constraining the range of possibilities subject to a complex set of influencing variables, quantitatively and/or qualitatively defined. The tolerance index is vital in order to model the uncertainty in the cost values within a realistic continuum as opposed to a single figure-of-merit. For this study, the tolerances, β, were adapted to follow those indicated by Ayyub (1997) and reported in Table 2

Table 2: Values of tolerance. Adapted from Ayyub (1997)

β	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Poor/Low	1.0	0.9	0.7	0.4	0	0	0	0	0	0	0
Median	0	0	0.4	0.7	0.9	1.0	0.9	0.7	0.4	0	0
High	0	0	0	0	0	0	0	0.4	0.7	0.9	0
Rather High	0	0	0	0	0	0.4	0.7	0.9	1.0	0.9	0.7

Each of the project variables in the validation set was converted into fuzzy set variables using Table 2

ANALYSIS AND DISCUSSION

Table 3 reports the performance of the NF hybrid models in predicting the final cost for 5 of the 99 different projects used in the validation set. The tolerance of each of the cost values in the validation set was computed using Eqn.4 and defuzzified to obtain a 3-point estimate representing the fuzzy mean, fuzzy upper and fuzzy lower values as illustrated in Table 4. These three values provided a range of likely final cost rather than the customary single value estimate. The overall results for the performance of the validation cases have been represented in Figure 3.

Table 3: Logarithmic Cost values for both composition operators

Project Validation cases	Max-Product Mean value	Max-min Mean Value	Actual Out-turn Cost value
Project Case 9	6.685	6.672	6.691
Project Case 204	5.592	5.572	5.670
Project Case 901	5.262	5.279	5.385
Project Case 505	5.877	5.934	5.980
Project Case 824	5.575	5.633	5.674

Based on statistical correlations, the max-product composition operator achieved on average a deviation of 1.71%; while the max-mean composition had an average deviation of 1.86%. The Max-Product composition performed consistently better in both the fuzzy mean and fuzzy lower values but did not show any significant advantage in the fuzzy upper cost values. This might indicate that the benefit of the max-product operator is situated within the fuzzy mean and lower cost target predictions.

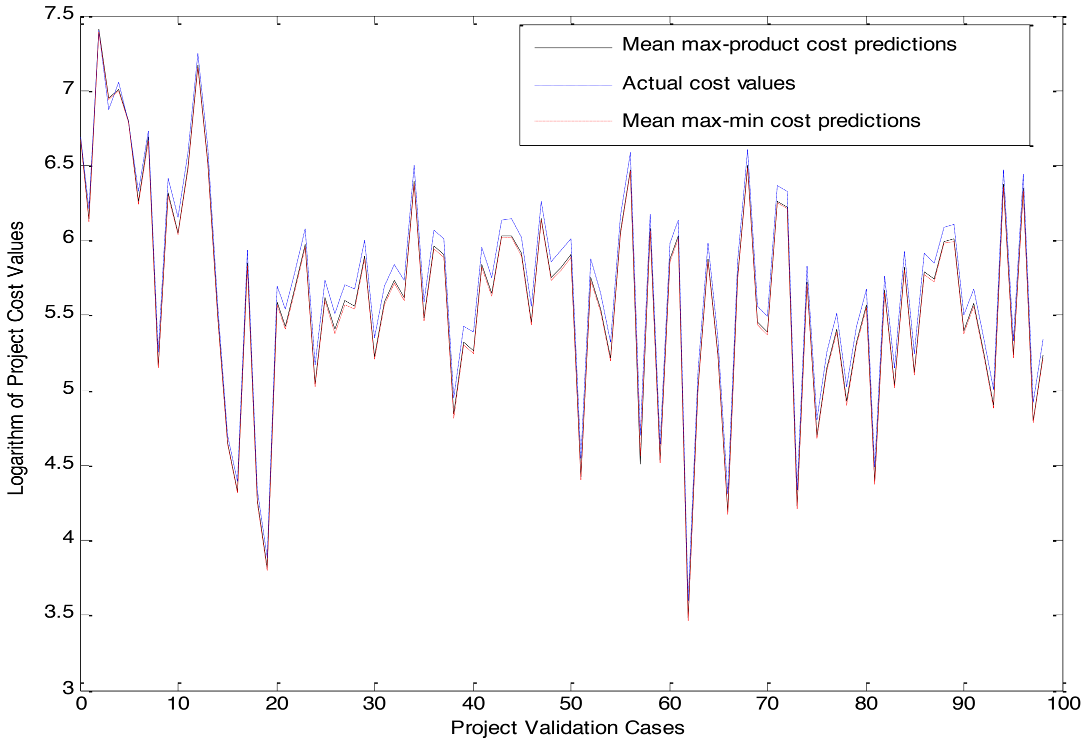

Figure 3 – Graphical plot of the project validation cases and the relational efficiency of composition operators

The corresponding percentage differences in the cost target were also estimated for all the 99 project validation cases. Table 4 provides a summary of the overall result obtained for all the validation cases.

Table 4: Summary of Results from Neuro-fuzzy Model Validation

Cost Category	Fuzzy Upper Value	Fuzzy Mean Value	Fuzzy Lower Value
Max-Min Operator	2.59%	2.07%	0.94%
Max-Product Operator	2.59%	1.74%	0.78%

The volatility measures considered for the range of values for the composition operators were fairly consistent. The standard deviation of the cost values of the max-product was £161,715, while that of the max-min was £188,506. This implies that the range of fluctuation in the max-min composition measure was higher than those obtained from the max-product composition predictions.

CONCLUSION

The research reported in this paper combines the learning and generalization capabilities of artificial neural networks with fuzzy logic's ability to formalise human reasoning and decision making within an environment of uncertainty and incomplete information. This paper develop a neuro-fuzzy hybrid cost model for predicting the final cost of small water infrastructure project and then evaluates the efficiency of the max-product and max-min composition operators in predicting the final target cost. Based on 99 project validation cases, it was found that the max-product composition operator achieved an average a deviation of 1.71% while the max-mean composition had an average deviation of 1.86%.

It is, however, noteworthy that these two composition operators are not an exhaustive treatment of the relational capabilities of fuzzy sets – they currently represent the most popular calculi employed in fuzzy set evaluations. There might be need to improve on the framework of the existing mathematical formulations of fuzzy sets in order to fully realize the potentials of fuzzy sets in modelling the vagueness in human reasoning and capturing irreducible uncertainties in water infrastructure projects. Improvements in the relational efficiency of neuro-fuzzy hybrid cost models will in no little way assist in developing a robust framework for realistic cost targets in water infrastructure projects.

REFERENCES

Ahiaga-Dagbui, DD and Smith, SD **(**2012) Neural networks for modelling the final target cost of water projects *In*: Smith, S.D (Ed) *Procs 28th Annual ARCOM Conference*, 3-5 September 2012, Edinburgh, UK, Association of Researchers in Construction Management, 307-316.

Ahiaga-Dagbui DD, Tokede O, Smith SD and Wamuziri S (2013) A neuro-fuzzy hybrid model for predicting final cost of water infrastructure projects *In:* Smith, S.D and Ahiaga-Dagbui, D.D (Eds) *Procs 29th Annual ARCOM Conference,* 2-4 September 2013, Reading, UK, Association of Researchers in Construction Management, 181-190.

Anderson, D. and G. McNeill (1992). "Artificial neural networks technology." A DACS (Data & Analysis Center for Software) State-of-the-Art Report, Contract Number F30602-89-C-0082: 87.

BCIS (2012). BIS Construction Price and Cost Indices. http://www.bcis.co.uk, Building Cost Information Services, UK.

Belohlavek, R. and G. J. Klir (2011). Concepts and fuzzy logic, MIT Press.

Boussabaine, A. and R. Kirkham (2008). Whole life-cycle costing: risk and risk responses, Wiley-Blackwell.

Boussabaine, H. and T. Elhag (1999). Tender Price Estimation Using ANN Methods, EPSRC Research Grant (GR/K/85001). Liverpool, UK, School of Architecture & Building Engineering, University of Liverpool.

Carpenter, G. A., S. Grossberg, et al. (1992). "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps." Neural Networks, IEEE Transactions on 3(5): 698-713.

Chan, A. P., D. W. Chan, et al. (2009). "Overview of the application of “fuzzy techniques” in construction management research." Journal of Construction Engineering and Management 135(11): 1241-1252.

Demicco, R. V. and G. J. Klir (2003). Fuzzy logic in geology, Academic Press.

Dubois, D. and H. M. Prade (2000). Fundamentals of fuzzy sets, Kluwer Academic Pub.

Eklund, P. (1994). "Network size versus preprocessing." Fuzzy Sets, Neural Networks and Soft Computing: 250-264.

Haykin, S. (1994). Neural networks: a comprehensive foundation, Prentice Hall PTR Upper Saddle River, NJ, USA.

Hegazy, T. (2002). Computer-based construction project management. Upper Saddle River, NJ, Prentice Hall Inc.

Hüllermeier, E. (1997) An approach to modelling and simulation of uncertain dynamical systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **5**(02), pp.117-137.

Kirkham, R. and P. S. Brandon (2007). Ferry and Brandon's Cost Planning of Buildings, John Wiley & Sons.

Kosko, B. and S. Isaka (1993). "Fuzzy logic." Scientific American 269(1): 62-67.

Lee, C. G. and C. Lin (1992). Supervised and unsupervised learning with fuzzy similarity for neural-network-based fuzzy logic control systems. Systems, Man and Cybernetics, 1992., IEEE International Conference on, IEEE.

Lee, S. H., F. Peña-Mora, et al. (2006). "Dynamic planning and control methodology for strategic and operational construction project management." Automation in construction 15(1): 84-97.

Loetamonphong, J. and S.-C. Fang (2001). "Optimization of fuzzy relation equations with max-product composition." Fuzzy Sets and Systems 118(3): 509-517.

Nicholas, J. M. (2004). Project management for business and engineering: Principles and practice. MA, USA; Oxford, UK, Elsevier Butterworth–Heinemann.

Pedrycz, W. (1984). "Identification in fuzzy systems." Systems, Man and Cybernetics, IEEE Transactions on(2): 361-366.

Pedrycz, W. (1996). Fuzzy modelling: paradigms and practice, Kluwer Academic Pub.

Ross, T. J. (2009). Fuzzy logic with engineering applications, Wiley.

Sanchez, E. (1976). "Resolution of composite fuzzy relation equations." Information and control 30(1): 38-48.

Smit, M. C. (2012). "A North Atlantic Treaty Organisation framework for life cycle costing." International Journal of Computer Integrated Manufacturing 25(4-5): 444-456.

Tokede, O. and S. Wamuziri (2012). Perceptions Of Fuzzy Set Theory In Construction Risk Analysis. Procs 28th Annual ARCOM Conference. S. Smith. Edinburgh. 2: 1197-1207.

Yager, R. R. and D. P. Filev (1994). "Essentials of fuzzy modeling and control." New York.

Zadeh, L. A. (2008). "Is there a need for fuzzy logic?" Information Sciences 178(13): 2751-2779.

Zimmermann, H. J. (2001). Fuzzy set theory-and its applications, Springer.